

ANALYSIS OF LINGUISTIC STEGANOGRAPHY TOOLS

A Thesis submitted to the Graduate Scholl in partial
fulfilment of the requirements for the degree
Master of Science (Information Technology),
Universiti Utara Malaysia

by

SHAIFIZAT BINTI MANSOR

Universiti Utara Malaysia
25th October, 2004

UNIVERSITI UTARA MALAYSIA
2004



JABATAN HAL EHWAL AKADEMIK
(Department of Academic Affairs)
Universiti Utara Malaysia

PERAKUAN KERJA KERTAS PROJEK
(Certificate of Project Paper)

Saya, yang bertandatangan, memperakukan bahawa
(I, the undersigned, certify that)

SHAIFIZAT BINTI MANSOR

calon untuk Ijazah
(candidate for the degree of) **MSc. (Information Technology)**

telah mengemukakan kertas projek yang bertajuk
(has presented his/her project paper of the following title)

ANALYSIS OF LINGUISTIC STEGANOGRAPHY TOOLS

seperti yang tercatat di muka surat tajuk dan kulit kertas projek
(as it appears on the title page and front cover of project paper)

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan.
(that the project paper acceptable in form and content, and that a satisfactory knowledge of the filed is covered by the project paper).

Nama Penyelia Utama
(Name of Main Supervisor): **MR. ROSHIDI DIN**

Tandatangan
(Signature)

:

Tarikh
(Date)

:

25/10/2004

PERMISSION TO USE

In presenting this thesis in partial fulfilment of the requirements for a post graduate degree from Universiti Utara Malaysia, I agree that Universiti Library may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by my supervisor(s) or, in their absence, by the Dean of Graduate School. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Request for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:

**Dean of Graduate School
Universiti Utara Malaysia
06010 UUM Sintok
Kedah Darul Aman**

ABSTRACT

It has been thought that the ability of sending a secret message through a network nowadays becomes more complicated and complex. One of the reasons why do we need tools to hide message is to keep conceal a secret message from being accessed by unauthorized and non-respective person. A technique such as steganography is one example method to send a secret message. There are several features encountered in steganography tools to ensure the tools run efficiently. This research tries to find the best fit available of linguistic steganography tools and evaluate the efficiency of tools based on the criteria selected. This research also tries to analyse several Linguistic Steganography that has been selected based on the software metric established.

ABSTRAK

Keupayaan menghantar mesej rahsia melalui sistem rangkaian kini semakin rumit dan kompleks. Kaedah-kaedah yang sesuai perlu dipilih untuk menghantar mesej bagi memastikan mesej yang dihantar hanya diketahui oleh pihak berkenaan. Teknik *steganography* yang telah diperkenalkan merupakan salah satu teknik yang digunakan untuk menghantar mesej rahsia melalui sistem rangkaian. Bagi menjamin keberkesanan penghantaran mesej dengan menggunakan kaedah tersebut, penilaian boleh dilakukan berdasarkan kepada beberapa kriteria yang boleh diukur. Kajian ini akan menilai teknik *steganography* yang boleh digunakan dengan efisien berdasarkan kepada beberapa kriteria yang telah ditetapkan. Kajian ini juga akan menganalisis beberapa teknik *steganography* yang telah dipilih berdasarkan kepada pengukuran pengaturcaraan yang digunakan.

ACKNOWLEDGEMENT

In The Name of Allah, The Most Gracious and The Most Merciful,

With Merciful and Blesses from Allah S.W.T, finally I am able to complete this final year project. I am very much grateful to the following list of individuals and organization for their helps and supports in helping me completing this research possible:

The Northern University of Malaysia for the facilities and resources provided,

My Supervisor, Mr Roshidi Din, for his helps and support in guiding me to conduct this study and during the writing of this thesis,

Northern University of Malaysia lecturers for their helps and lessons contributed throughout the completion session of MSc IT,

My friends and others for their support and expertise in building the applications,

And special thanks to my parents and my family for their support and understanding.

SHAIFIZAT BINTI MANSOR
Graduate School
Universiti Utara Malaysia
October, 2004

TABLE OF CONTENTS

PERMISSION TO USE	I
ABSTRACT	II
ABSTRAK	III
ACKNOWLEDGEMENT	IV
TABLE OF CONTENTS	V
LIST OF TABLE	VII
LIST OF FIGURES.....	VIII
CHAPTER ONE	1
INTRODUCTION.....	1
1.1 Problem Statement	3
1.2 Project's Objective	4
1.3 Scope of Research	4
1.3.1 Linguistic Steganography.....	5
1.3.2 Software Metric.....	5
1.3.3 Size of File	6
1.4 Significance of Study	6
1.4.1 Analysis of Linguistic Steganography	6
1.4.2 Towards Better Performance.....	6
1.5 Organization of the Thesis	7
CHAPTER TWO	9
LITERATURE REVIEW	9
2.1 Introduction	9
2.2 Linguistic Steganography.....	9
2.3 Software Metric.....	10
2.4 Tools Used	12
CHAPTER THREE.....	13
RESEARCH METHODOLOGY	13
3.1 Background Study and Literature Review	14
3.2 Compare the Linguistic Steganography Tools.....	14

3.3	Analysis of Steganography Tools	15
3.3.1	Software Metric.....	15
3.3.2	Execution Time	17
3.3.3	Coding Segments	17
3.3.4	Testing File.....	17
CHAPTER FOUR.....		19
DATA ANALYSIS.....		19
4.1	Selected Linguistic Steganography Tools.....	19
4.2	Analysis of Linguistic Steganography Tools	21
4.2.1	Metric Summary.....	21
4.2.2	Execution Time	25
4.3	Result.....	30
CHAPTER FIVE.....		33
DISCUSSION		33
5.1	Metric Summary.....	33
5.2	Execution Time	35
CHAPTER SIX		37
CONCLUSION		37
REFERENCES.....		38
APPENDICES.....		41
APPENDIX A		42
APPENDIX B		48
APPENDIX C		61
APPENDIX D.....		72
APPENDIX E.....		78
APPENDIX F.....		86

LIST OF TABLE

Table 3.1: Distribution of File Size Category	18
Table 4.1: Summary Metric of Linguistic Steganography Tools.....	25
Table 4.2: An execution time of every Linguistic steganography tool	30

LIST OF FIGURES

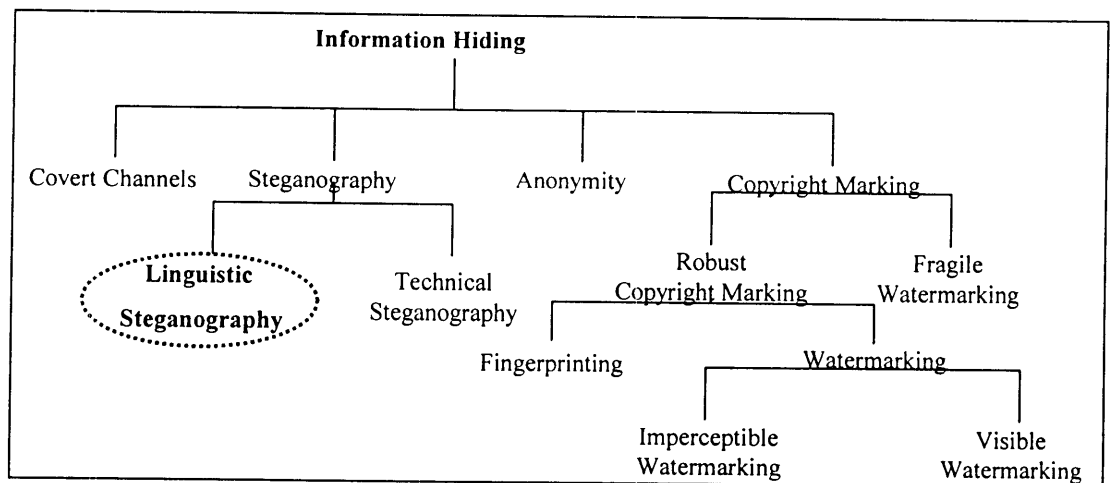
Figure 1.1: Linguistic Steganography in Information Hiding Domain.....	1
Figure 1.2: Process of Encoding Message Using Steganography Technique.....	2
Figure 3.1: Research Methodology Steps	13
Figure 3.2: Tested Tool	14
Figure 3.3: Program Segments.....	17
Figure 4.1: Percent Lines with Comments of Steganography Tools by Percentage..	22
Figure 4.2: Average Statements per Function of Steganography Tools by Average.	23
Figure 4.3: Average Block Depth of Steganography Tools by Depth of Block	24
Figure 4.4:PGP Keytools	27
Figure 4.5: Execution Time for Steganography Tools of 10 bytes Character.....	30
Figure 4.6: Execution Time for Steganography Tools of 100 bytes Character	31
Figure 4.7: Execution Time for Steganography Tools of 1000 bytes Character	31
Figure 4.8: Execution Time for Steganography Tools of 20000 bytes Character	32
Figure 5.1: Linguistic Steganography Tools Average Statements per Function and Average Block Depth.....	34
Figure 5.2: Linguistic Steganography Tools Percent Lines with Comments.....	35
Figure 5.3: Execution Time of Steganography Tools Overall Result by Time.....	36

CHAPTER ONE

INTRODUCTION

With the advanced technology, the ability of sending secret messages becomes more complex and complicated. There are tools developed to ensure the secrecy of the message whenever data is sent from sender and receiver. How secure the data is based on the efficiency of tools performance. Steganography, one form of information hiding can be divided into two general categories, which are technical and linguistic steganography. Linguistic steganography is any form of steganography where the carrier media is text, WbStego (2004). Figure 1.1 shows the diagram of information hiding.

Figure 1.1: Linguistic Steganography in Information Hiding Domain



The contents of
the thesis is for
internal user
only

REFERENCES

- Anderson, R. J., and Petitcolas F. A., (1998). On the Limit of Steganography. *IEEE Journal of Selected Areas in Communications*, 16(4), 474-481.
- Artz, D., (2001). Digital Steganography: Hiding data within Data.. *IEEE Internet Computing*, 75 – 80.
- Duncan, A. S (1988). Software Development Productivity Tools and Metrics. *IEEE Journal*, 41 – 48.
- Flowers, D. A. (2000). Investigating Steganography. *ACM Southeast Regional Conference* 468 – 469.
- Hetzl, S. (2003). Steghide – *Manual*. Retrieved July 20 2004 from <http://steghide.sourceforge.net>
- Johnson, A. M. J, & Malek, M. (1998). Survey of Software Tools for Evaluating Reliability, Availability and Serviceability. *ACM Computing Survey*, 20(4), 227 - 269.
- Kafura, D. & Canning, J. (1985). A validation of Software Metrics Using Many Metrics and Two Resources. *IEEE Journal*.
- Kwan, M. (2001). *SNOW Software*. Retrieved June 20 2004 from <http://www.darkside.com/au>
- Komoski, P.K. (1995). Seven Steps to Responsible Software Selection. Eric Digest. *Information Analysis*.
- Lee, Y. K., & Chen, L. H. (2000). A Secure Robust Image Steganography. Retrieved May 20 2004 from <http://www.citeseer.org>

- Loosemore, S., McGrath, R., Oram, A. & Stallman, R.M. (1993). *Processor Time*. Retrieved August 28 2004 from http://www.cs.utah.edu/dept/old/textinfor/glibr_manual_0.02/library.html
- Maher, K. (2001). *TEXTO*. Retrieved July 28 2004 from <http://www.eberl.net/texto>
- Mathias, K.S, Cross, J.H., Hendrix, T.D & Barowski, L. R, (1999). The Role of Software Measure and Metrics in Studies of Program Comprehension. *Journal of ACM*, 1-58113-128-3
- Moskowitz, I. S., Chang, L. W., & Newman, R. E. (2002). Securing Information: Capacity is the Wrong Paradigm. *New Security Paradigm Workshop*, 114-126.
- Provos, N. & Honeyman, P. (2003). Hide and Seek. *IEEE Department Security and Privacy*, 1(3), 32-44.
- Nystedt, S. & Sandros, C. (1999). *Software Complexity and Project Performance*. Master Thesis : 365-369.
- Petitcolas, F.A.P, Anderson, R. J & Kuhn, M. G (1999). Information Hiding – A Survey, *IEEE Computers*, 87(7), 1062 –1078.
- Petitcolas, F.A.P (2000). *Watermarking Schemes Evaluation*. Retrieved June 20 2004, from <http://www.petitcolas.net/fabien/publications/ieeespm00-evaluation.doc>
- Ochiai, K., Iwasaki, H., Naganuma, J., Endo, M., & Ogura, T. (1999). High-Speed Software-Based Platform for Embedded Software of A Single-Chip MPEG-2 Video Encoder LSI With HDTV Scalability. *Proceeding of the Conference on Design, Automation and Test in Europe*: 1-6.
- Ranum, D. L. (1998). *Teaching and Learning in Computer Science Incorporating Assessment as a Tool Against Obsolescence*. Retrieved August 20, 2003 from <http://www.pkal.org/template2.cfm>
- Satyanarayanan, M. (1981). A Study of File Sizes and Functional Lifetimes. *Symposium on Operating System Principle.*: 96 – 108.

- Schroder, B (1998). *Pretty Good Envelope*. Retrieved June 20 20004 from <http://www.burks.de/stegano/pge.html>
- Scotto, M., Sillitti, A., Succi, G & Vernazza, T (2004). A relational Approach to Software Metrics. *Journal of ACM*, 1(2), 1536 – 1540.
- Welzel, D., & Hausen, H. L. (1993). A Five Steps Method for Metric-Based Software Evaluation: Effective Softwrae Metrication with Respect to Quality Standard. *Journal of ACM*, 39(2–5), 273 – 276.
- Westfield, A. & Pfitzman, A. (2000). *Attacks on Steganography Systems*. Retrieved June 20 2004 , from <http://www.citeseer.org>
- Wikipedia (2004). *The Free Encyclopedia*. Retrieved August 27 2004. from http://en.Wikipedia.org/wiki/Software_metric
- Wanner, J. F. (2001). *Campwood Software*. Retrieved August 20 2004 from <http://wwwcapmwoodsw.com/cwservices.html>
- Zuse, H. & Bollmann, P. (1998). Software Metric. *SIGPLAN Notices*, 24 (8), 23 - 33.